

### **Amendments**

#### **In the Specification:**

Please replace the section entitled "Brief Description of the Drawings, located on pages 7-16, with the following:

Figure 1 is a schematic of a vector useful for the invention. In this example, integration of a marker peptide coding sequence can occur either in an intron or exon in split genes encoding protein products (inclusive but not limited, e.g. genes without introns that encode proteins such as histones etc., or genes encoding physiologically active RNAs, eg., snRNA, scRNA, spliceosome components etc.). For the sake of clarity, integration into an intron sequence of a cellular gene encoding a protein is shown. Placement of a splicing acceptor (SA) upstream of a marker peptide-encoding sequence results in the synthesis of a mRNA encoding a fusion protein that includes the marker peptide fused to peptide sequences encoded by upstream exons (occurs when the splice donor of the nearest upstream exon (closer to the start of transcription) is reacted to the splice donor present in the integrated marker DNA sequence).

Figures 2A-F depict diagrams of several variant constructions of retroviral vectors which perform certain distinct functions for acquiring different types of information in cells. The critical portion is the area located between the 5' and 3' LTR. These expression cassettes would be moved essentially intact between any of the various viruses and/or plasmids that we have mentioned. Figure 2A depicts a vector for exon acquisition. Figure 2B depicts a vector designed for integration site acquisition. Figure 2C depicts a vector for incorporation of multiple marker genes. Figure 2D depicts a transfection cassette. Figure 2E depicts a vector for replication competent virus. Figure 2F depicts a vector for fusion protein marker for cell pre-separation and

FACS analysis. RE Type IIS restriction enzyme site; LTR, long terminal repeat; CMV IE, CMV intermediate early promoter; NeoR: neomycin resistant gene; pA, bovine growth hormone poly-A signal; SA, human gamma-globin intron #2 splicing acceptor; pA, NeoR, CMV, hrGFP, SA are in anti-sense orientation against LTRs. Gag, pol, env, retroviral helper virus.

Figure 3 delivers a rudimentary overview of the process of the invention. The process begins with two different populations of cells to be compared. Each population of cells to be compared will have been marked genetically by a vector containing marker/s-peptides to facilitate detection and determination of relative concentration of marker/s. Left portion of middle panel demonstrates separation of populations of cells based on relative amount of marker present in the tagged cells. Sequences flanking the vector will be determined by but not limited to serial analysis of viral integration (SAVI) or sequence tag acquisition and reporting system (STARS) methods. Valid tags will then be compared to public and commercial data bases and annotated into our own data bases.

Figures 4A and B depict a gene trap vector, pGT5A with a humanized rellina fluorescence protein (hrGFP) as an assay marker, or reporter gene. Figure 4A is a schematic diagram of pGT5A plasmid. LTR, long terminal repeat; PBS, retroviral primer binding site; CMV IE, CMV intermediate early promoter; NeoR: neomycin resistant gene; pA, bovine growth hormone poly-A signal; SA, human  $\gamma$ -globin intron #2 splicing acceptor; AmpR, ampicillin-resistant gene for bacterial cloning. pA, NeoR, CMV, hrGFP, SA are in anti-sense orientation against LTRs. Figure 4B is a schematic of the order of genes in pGT5A vector.

Figures 5A and B depict a vector, pGT5AH with a humanized rellina fluorescence protein (hrGFP) as an assay marker, or reporter gene. Figure 5A is a schematic diagram of pGT5AH plasmid. LTR, long terminal repeat; PBS, retroviral primer binding site; CMV IE,

CMV intermediate early promoter; NeoR, neomycin resistant gene; pA, bovine growth hormone poly-A signal; SA, human  $\gamma$ -globin intron #2 splicing acceptor; AmpR, ampicillin-resistant gene for bacterial cloning. pA, NeoR, CMV, hrGFP, SA are in anti-sense orientation against LTRs. His6 tag contains 6 continuous histidine residue at c-terminal of hrGFP for detection by anti-His6 antibody. Figure 5B is a schematic of the order of genes in pGT5AH vector.

Figures 6A and B depict pGT5Z with a humanized rhesus fluorescence protein (hrGFP) as an assay marker, or reporter gene and Zeocin-resistance gene (ZeoR). Figure 6A is a schematic diagram of pGT5Z plasmid. LTR, long terminal repeat; PBS, retroviral primer binding site; CMV IE, CMV intermediate early promoter; NeoR, neomycin resistant gene; pA, bovine growth hormone poly-A signal; SA, human  $\gamma$ -globin intron #2 splicing acceptor; SD, synthetic splicing donor. SV40, simian virus type 40 early promoter. AmpR, ampicillin-resistant gene for bacterial cloning. pA, NeoR, CMV, hrGFP, SA are in anti-sense orientation against LTRs. Figure 6B is a schematic of the order of genes in pGT5Z vector.

Figures 7A and B depict a demonstration of the splicing function and fusion hrGFP protein expressed by pGT5A vector. Figure 7A depicts a construct of pGT5Z, which is derived from pGT5A with an insertion of a SV40 early promoter (SV40), Zeocin-resistant gene (ZeoR), and a synthetic splicing donor and partial intron to demonstrate the expected biological functions of pGT5A after gene trapping. Figure 7B demonstrates that pGT5Z-transfected cells after Zeocin selection showed significant Zeocin-hrGFP fusion protein expression by FACS analysis.

Figures 8A and B depict a gene trapping of PGT5A-transfected PA317 cells. Figure 8A demonstrates that PA317 cells transfected with pGT5A showed a 3.6% of hrGFP-positive cell population. Figure 8B demonstrates that sorting of the hrGFP-positive cell population in Figure

8A by FACS cell sorter. hrGFP-positive population was enriched to 95% after 2 weeks of cell culture.

Figure 9 is a depiction of gene expression of hrGFP in gene trapped PA317 cells. RT-PCR was performed on total RNA extracted from sorted cells in Figure 7 and Figure 8, and PCR product was electrophoresed in 2% agarose gel. The whole length of hrGFP transcripts driven by trapped cellular promoter (GT5A/PA317) were amplified by hrGFP specific primers after cDNA synthesis as indicated with an arrow. Transcripts from GT5Z in PA317 (GT5Z/PA317) and PA317 without vector (PA317) were used as a positive and negative control.

Figures 10A and B depict gene trapping of GT5A vector in human lung cancer cells. A549, after viral transduction. Figure 10A demonstrates A549 cells without transduction analyzed by FRCS. Figure 10B demonstrates that A549 cells with GT5A- transduction analyzed by FACS showed the hrGFP-positive population is 1.68% after gene trapping.

Figure 11 is a depiction of gene trapping of GT5A vector in NIH3T3 cells. Mixed population of GT5A-trapped NIH3T3 cells were sorted and cultured for three weeks and then analyzed by FACS comparing to untransduced cells. Different intensities of hrGFP were shown in four different major groups.

Figure 12 is a depiction of hrGFP gene expression of single-cell clones from GT5A-trapped NIH3T3 cells. Individual single-cells were sorted into 96-wells plate and cultured to a sufficient population for FACS analysis. A6P1 and C4P2, C8P2 and H8P2 were analyzed at two different events while compared to untransduced NIH3T3 cells.

Figures 13A-D depict gene trapping with an  $\alpha$ 1.3-galactosyl transferase as a reporter gene in human melanoma cell line, A375. Figure 13A is a schematic diagram of serial gene trapping vectors with ( $\alpha$ 1.3-galactosyl transferase ( $\alpha$ 1.3-gal) gene. LTR, long terminal repeat; SV40.

simian virus type40 early promoter; ZeoR, Zeocin resistant gene; CMV, CMV early promoter; NeoR; neomycin resistant gene; pA, bovine growth hormone poly-A signal. SA, human  $\gamma$ -globin intron2 splicing acceptor; SD, synthetic splicing donor. pA, NeoR, CMV,  $\alpha$ 1.3 gal, SA or SD. ZeoR and SV40 are in anti-sense orientation against LTRs. Figure 13B demonstrates gene trapping of pGT7A in A375/AMIZ cells. Cells were labeled with lectin conjugated with FITC for FACS analysis. Lectin binds to  $\alpha$ 1.3 gal epitopes on cell surface to show successful gene-trapping. Figure 13C demonstrates gene trapping in A375/AMIZ cells 3 days post transfection of pGT7AH. Figure 13D demonstrates that splicing function and functional  $\alpha$ -1.3  $\alpha$ -gal/ZeoR fusion protein were demonstrated by lectin/FITC-positive cells.

Figure 14 is a schematic depicting a vector of the invention which utilizes homologous recombination as the integration strategy. The repeat sequences are engineered to flank the assay marker gene and then introduced to the cell.

Figure 15 is a diagram depicting the concept of frame alignment. Only 1 in 3 integrants will be in frame, based upon the triplet codon scheme so that only 1 in three integrated vectors will be functional and result in translation of the assay marker.

Figure 16 is a schematic depicting the STARS process. A method of cleaving said cellular DNA such that inserted DNA (with sequence known to the operator) is cleaved once and flanking cellular DNA of unknown sequence is cleaved again in the regions contiguous to the inserted piece of DNA. Cleavage of the DNA occurs in a fashion generating ends that permit the circularization of DNA fragments producing a molecule with the sequence known to the operator flanking both sides, and continuous with, a variable length of cellular DNA of unknown sequence. The region containing the unknown DNA is then amplified and sequenced.

Figures 17A and B depict the SAVI process. Integration of a marker gene can occur either in an intron or exon. Adjacent a splicing acceptor (SA) in front of a marker gene can therefore result in a fusion protein for marker gene expression after the integrated gene exon region is spliced into the SA signal of the marker gene. However, to sequence the exon region of this integrated gene to release the identity becomes a problem.

To overcome this obstacle, a Type IIS restriction enzyme (RE) recognition site will be introduced between the SA signal and the start codon (ATG) of marker genes, such as hrGFP, alpha 1-3 galactosyltransferase ( $\alpha$ -gal), etc. This can be illustrated as SA-RE-ATG. This RE site can be designed in frame with markers. After the SA joins to the splicing donor (SD) of the integrated cellular gene by cellular splicing mechanism, reverse transcription will be employed to convert this hybrid RNA transcript into a complementary DNA (cDNA) (inclusive of, but not limited to, cDNA as cellular DNA may be used). This cDNA will then be subjected to RE digestion of exon from the integrated gene ten to twenty bases away from the SD/SA depending on which RE is used. A biotin-labeled primer #1 designed for a known marker (MK) gene is then employed to extend the ssDNA into this exon. Collection of this biotin-ssDNA by streptavidin conjugated magnetic beads will enrich these specific ssDNA for DNA terminal transferase reaction. Polymer deoxynucleotide can be added onto these ssDNA as a tail at their 3' end. A polymer primer complementary to the polymer tail and a second primer #2 on MK marker gene can therefore be used to amplify this 3' end of exon region. These short tags from different integrated genes by ligation reactions into a longer DNA fragment that is subsequently sequenced. Sequencing results of these tags can be used to retrieve the identity from EST databases or genomic databases. This approach can utilize all possible gene transfer methods to deliver above construct into DNA or RNA genomes of all organisms.

Figures 18A and B depict a non-limiting flow diagram demonstrating the entire process. This figure delivers a rudimentary overview of the process of the invention. The process begins with two different populations of cells to be compared. Each population of cells to be compared will have been marked genetically by a vector containing marker/s-peptides to facilitate detection and determination of relative concentration of marker/s. Left portion of middle panel demonstrates separation of populations of cells based on relative amount of marker present in the tagged cells. Sequences flanking the vector will be determined by but not limited to serial analysis of viral integration (SAVI) or sequence tag acquisition and reporting system (STARS) methods. Valid tags will then be compared to public and commercial data bases and annotated into our own data bases. As can be seen at each stage alternatives exist for each step.

Figure 19 is a diagram demonstrating the layers of information which may be assayed to identify the real state of cell (furthest outward circle). Those who assay DNA and raw sequence data determine gene function based on sequence similarity, gene structure, and evolutionary relationships. Missing from this data is any mRNA or translational modification data. Those who assay mRNA gain a prediction of a protein profile based on the assumption that protein levels are directly proportional to mRNA. An assumption which is proving to be erroneous. Closest of all these methods to the real cell state is the method of the invention which detects actual cellular protein levels by direct measurement.

Figure 20 is a depiction of a successful gene trapping in pGT5A-transfected PA317 cells. NcoI restriction site located at the 5' end of hrGFP marker gene and an EcoRI at the Oligo-dA primer were used as cloning sites for gene trapped sequence into a sequencing vector which was digested with NcoI and EcoRI. After BLAST searching against mouse EST database in GenBank, the sequence trapped by pGT5A demonstrates 99% homology to a high mobility

group protein, HMGI-C, a nuclear phosphoprotein that contains three short DNA-binding domains (AT-hooks) and a highly acidic C-terminus.

Interest in this protein has recently been stimulated by three observations: the expression of the gene is cell-cycle regulated, the gene is rearranged in a number of tumors of mesenchymal origin and mice that have both HMGI-C alleles disrupted exhibit the pygmy phenotype. These observations suggest a role for HMGI-C in cell growth, more specifically, during fetal growth since the protein is normally only expressed in embryonic tissues. It is likely that the HMGI-C protein acts as an architectural transcription factor, regulating the expression of one or more genes that control embryonic cell growth. Since HMGI-C binds to the minor groove at AT-rich DNA this interaction could be a target for minor groove chemotherapeutic agents in the treatment of sarcomas expressing a rearranged gene.

Figure 21 is a depiction of gene trapping of an exon with unknown biological function in pGT5A-transfected PA317 cells. NcoI restriction site located at the 5' end of hrGFP marker gene and an EcoRI at the oligo-dA primer were used as cloning sites for gene trapped sequence into a sequencing vector which was digested with NcoI and EcoRI. After BLAST searching against the EST database in GenBank, the sequence trapped by pGT5A is 95% match to a NCI\_CGAP\_Li9 Mus musculus cDNA clones, BF539247.1/BF533319.1/...etc., which have been found in the cDNA libraries from Salivary gland and liver.

---

Please replace the paragraph beginning at page 31, line 6, and extending to page 32, line 6, with the following:

---

B2 Unless otherwise stated, sequence identity/similarity values provided herein refer to the value obtained using the BLAST 2.0 suite of programs using default parameters. Altschul et al., *Nucleic Acids Res.* 25:3389-3402 (1997). Software for performing BLAST analyses is publicly



2  
available, e.g., through the National Center for Biotechnology-Information. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always  $> 0$ ) and N (penalty score for mismatching residues; always  $< 0$ ). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a word length (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a word length (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915).

**In the Claims**

Please cancel claims 1-44, 51, and 52.

Please add the following claims: